

Databasebenchmarks

Kristian Dahl, Heidi Gregersen, Christer A. Have, Christian S. Jensen, Janne Sigurðsson, Janne S. Winter.

Databasesystemer anvendes i dag som standard-komponenter i de fleste større administrative edb-systemer. Da antallet af database-produkter på markedet er voksende, ligger der en stigende udfordring i at vælge det databasesystem, der er bedst egnet til brug i et konkret edb-system.

Der stilles ofte performancekrav til edb-systemer, og performance er da også en væsentlig konkurrence-parameter. Samtidig varetager databasesystem-komponenten af et edb-system ofte tidskrævende operationer og er en flaskehals. Dette gør performance-egenskaber til en væsentlig faktor ved valg af databasesystem.

Databasebenchmarks er en betegnelse for de værktøjer, der anvendes til sammenligning af databasesystemers performance-egenskaber. En benchmark er defineret og beskrevet i en benchmarkspecifikation. En sådan specifikation definerer en testdatabase og en mængde af forespørgsler. Den rummer desuden en testprocedure, der beskriver under hvilke betingelser, forespørgslerne skal udføres. Testproceduren beskriver også hvilke delresultater, der indsamles under udførelsen af testen, og hvordan det samlede resultat af benchmarktesten beregnes.

Benchmarktyper

Der findes mange forskellige typer af benchmarks. Disse kan klassificeres efter den teknologi, databasesystemet bygger på, samt den anvendelsessituation, hvori systemet ønskes benyttet.

Visse benchmarks tester således alene relationelle databasesystemer, hvori- mod andre tester eksempelvis objektorienterede databasesystemer. Der er så store forskelle mellem disse slags databasesystemer, både med hensyn til opbygning og de anvendelsesområder, de bedst egner sig til, at det er urealistisk at lave en enkelt benchmark, der dækker begge slags systemer. Benchmarks er med andre ord *teknologispecifikke*.

For hver slags databasesystem kan der laves benchmarks efter to forskellige strategier. De kan laves som anvendelsesspecifikke benchmarks eller som generelle benchmarks. Afhængig af den påtænkte anvendelse af et databasesystem har hver strategi fordele og ulemper.

En *anvendelsesspecifik benchmark* er rettet mod én bestemt type af anvendelsessituationer, såsom enten bankvirksomhed, salgs- og distributionsvirksomhed eller statistisk databehandling. Sådanne benchmarks tilstræber at skabe en test-

situation med en database og forespørgsler, der er realistiske for den givne anvendelsestype. En anvendelsesspecifik benchmark er klart at anbefale, vel at mærke hvis en sådan findes.

Hvis der ikke findes en anvendelsesspecifik benchmark svarende til en given anvendelsessituation, kan man i stedet anvende en generel benchmark. *Generelle benchmarks* giver et mindre detaljeret, bredere billede af et systems performance-egenskaber, idet de tester et bredt udvalg af basale, generelle systemegenskaber, herunder eksempelvis de såkaldte ACID-egenskaber. Hvis anvendelsen af et databasesystem forventes at ændre sig væsentligt, er det specielt fordelagtigt at anvende en generel benchmark som basis for performancevurdering af databasesystemer.

Benchmarkkomponenter

For at forstå i større detalje hvad en benchmark omfatter, er det på sin plads at se på de fire benchmarkkomponenter: Testdatabasen, forespørgslerne, testproceduren og testresultatet.

Testdatabasen, som forespørgslerne skal udføres på, kan i nogle tilfælde konstrueres ved brug af en databasegenerator. I andre tilfælde er det blot nøje beskrevet, hvorledes testdatabasens skemaer skal se ud, og på hvilken måde data skal genereres og fordeles i databasen.

Alt efter hvilken type benchmark der er tale om, er der i benchmarkspecifikationen eller den tilhørende software angivet et antal *forespørgsler*, der skal benyttes til selve testen. Disse forespørgsler skal hver for sig være realistiske og skal optræde i et realistisk miks. De skal med andre ord afspejle en reel anvendelse af databasesystemet. Dette gælder specielt, såfremt der er tale om en anvendelsesspecifik benchmark.

En *testprocedure* er en beskrivelse af, hvad der skal testes, og på hvilken måde selve benchmarktesten skal udføres. Der angives således i hvilken rækkefølge og under hvilke omstændigheder, de enkelte forespørgsler skal udføres. Nogle testprocedurer indeholder både en- og flerbrugertests, mens andre kun indbefatter den ene test. En testprocedure kan også indeholde tests af diverse kontrolmekanismer (herunder recovery og concurrency control (ACID-egenskaber)), der findes i databasesystemer.

En idéel testprocedure sikrer *gentagelighed*, altså at man får det samme benchmarkresultat, når benchmarken udføres gentagne gange. Gentagelighed er et essentielt krav til benchmarks, da det giver mulighed for at verificere benchmarkresultater. Dét, at forkerte resultater kan afsløres, er med til at give mere pålidelige resultater.

Resultatet af at udføre en benchmark sammensættes af tidsmålinger for hver operation, der er udført under testen. En benchmarkspecifikation angiver, hvordan disse delmålinger skal indgå i det endelige testresultat, som kan have vidt forskellige former.

Et benchmarkresultat kan udtrykkes som forespørgslernes eksekveringshastigheder set i forhold til de omkostninger, der er forbundet med køb af systemet. Omkostningerne omfatter ofte al nødvendig hard- og software samt vedligeholdelse af dette, f.eks. over en fem-årig periode. Nogle benchmarks producerer endvidere testresultater for flere forskellige databasestørrelser. Dermed kan man anvende det resultat, der svarer til den databasestørrelse, der er mest relevant for éns situation.

Andre benchmarks har et resultat, der udtrykker, hvor stor testdatabasen kan være, når de foreskrevne forespørgsler skal kunne udføres inden for et fastsat tidsrum. Endelig inkluderer mange testresultater — foruden en af de ovenstående målinger — også mere detaljerede målinger såsom antallet af I/O-tilgange under forespørgslernes udførsel samt anvendt CPU-tid.

Fortolkning af benchmarkresultater

Det er ikke altid nødvendigt selv at udføre en benchmarktest for at fremskaffe benchmarkresultater. Visse benchmarkudviklere har en central styring af benchmarkresultater. Dette har den fordel, at det er muligt at fremskaffe benchmarkresultater ved at kontakte benchmarkudviklerne. En anden mulighed er at fremskaffe benchmarkresultater fra fagblade, der fra tid til anden vurderer databasesystemer og publicerer resultaterne af undersøgelserne.

Når man baserer sig på eksisterende benchmarkresultater, skal man, for at undgå at misfortolke og misbruge benchmarkresultaterne, være opmærksom på visse faktorer. Det er vigtigt, at det anvendelsesområde, som benchmarken dækker, svarer til den anvendelsessituation, som databasesystemet ønskes benyttet til. Er dette ikke tilfældet, kan benchmarkens resultater være yderst misvisende.

Opmærksomheden skal også være rettet mod, at de fleste benchmarks på markedet har til formål at teste selve databasesystemet. Resultater fra sådanne benchmarks anvendt på forskellige databasesystemer kan være yderst misvisende og direkte usammenlignelige, hvis testene er udført på forskellige platforme. Dette skyldes, at eksempelvis en kraftigere CPU kan forbedre benchmarkresultaterne betydeligt. Man opnår en mere troværdig angrebssvinkel, når det er benchmarkens mål at evaluere performance for et fuldt konfigureret system, der således indbefatter hardware og operativsystem samt anden nødvendig software.

De ovennævnte faktorer omhandler korrekt anvendelse af benchmarks. Men for at benchmarkresultater skal kunne betragtes som troværdige, skal selve benchmarken desuden opfylde en række krav. For testdatabasen er dette eksempelvis realisme af dens design og af fordelingen af de enkelte attributters værdier samt mulighed for at skalere testdatabasens størrelse, således at denne kan tilpasses anvendelsesområdet.

Op gennem 70'erne og i begyndelsen af 80'erne blev der set flere eksempler på misbrug af benchmarks. Den form for misbrug, der her fandt sted, benævnes benchmarkkrige og benchmarketing. *Benchmarkkrige* startede, når en virksomhed, der udviklede databasesystemer, miskrediterede benchmarkresultater fra konkurrerende databasesystemer. De forsøgte dernæst at overgå konkurrentens resultater ved at foretage småjusteringer af deres eget databasesystem. *Benchmarketing* tager form af bevidst konstruktion af en benchmark, der sætter virksomhedens eget databasesystem i godt lys. Dette er muligt ved at begrænse benchmarken til kun at teste de dele af systemet, der fungerer bedst.

Alternative vurderingsmetoder

Det er muligt at vurdere databasesystemers performance ved at konstruere og dernæst benytte en egen test eller ved at kombinere brugen af benchmarks og egen test. Fra en systemkøbers synsvinkel har to faktorer indflydelse på valget: Tilstedeværelse af specielle krav til databasesystemet og eksistens af en acceptabel benchmark inden for den påtænkte anvendelse.

Hvis der fra køberens side er specielle performance- eller kapacitetskrav, kan det enten diktere valget af databasesystem, eller det kan betyde, at det er nødvendigt at konstruere sin egen test med henblik på at teste databasesystemet i forhold til de specielle krav. Denne test kan enten laves gennemgribende og dermed benyttes som eneste valggrundlag, eller den kan laves simpel og kombineres med brug af benchmarks.

Hvis der ikke eksisterer en acceptabel benchmark til test af det anvendelsesområde, hvor databasesystemet skal bruges, er det nødvendigt med en gennemgribende egen test af systemet — forudsat at køberen *vil* have en performancetest som valggrundlag.

Med hensyn til fordele og ulemper for diverse vurderingsmetoder gælder, at både udførsel af en kendt benchmarktest og specielt konstruktion af egen test har den ulempe, at de er yderst ressourcekrævende. Derimod har konstruktion af egen test den fordel, at det er muligt at afpasse testen efter egne ønsker i forhold til den konkrete, påtænkte anvendelse. Ved brug af eksisterende benchmarkresultater kan det være en stor fordel at vælge en meget anvendt

benchmark, da der dermed findes flere testresultater. Dermed er der automatisk flere sammenligningsmuligheder.

Råd og vink ved udvikling af egen test

Når en virksomhed skal udvikle sin egen test, bør opmærksomheden rettes specielt mod de følgende fire aspekter:

- For at testen er *relevant for anvendelsesområdet*, skal det sikres, at testdatabasen og forespørgslerne, herunder deres blandingsforhold og rækkefølge, er realistiske.
- Hvis testen indeholder en flerbrugertest, bør denne omfatte det *antal brugere, der normalt vil optræde i anvendelsessituationen*. Gennem simulering af brugernes aktivitet kan man opnå det mest realistiske billede af systemets performance. Flerbrugertesten bør ligeledes omfatte test af relevante concurrency control og recovery mekanismer.
- Det skal tilstræbes, at testen fremstår så *simpel* som muligt, hvilket kan opnås ved nøje at gennemtænke et design for testen, inden den implementeres. Det anbefales dog samtidig, at simpliciteten ikke opprioriteres på bekostning af testens relevans.
- I testen bør der ikke benyttes hardware-specifikke krav, idet testen så vil være besværlig at gennemføre på anden hardware. *Portabilitet* er af afgørende betydning ved performancetests, idet samspillet mellem hardware og databasesystem har stor betydning for performance.

Videre læsning

Denne oversigtsartikel er skrevet på baggrund af rapporten "Vurdering af benchmarks som redskab til evaluering af DBMS-performance," udarbejdet af forfatterne i efteråret 1993. Rapporten opstiller idealer, som bør opfyldes af en benchmarkspecifikation, og vurderer i hvor høj grad tre konkrete benchmarks opfylder disse idealer. Desuden peges der på modsætninger mellem de enkelte idealer. Man kan for 100 kr. erhverve en kopi af rapporten, som er på 155 sider, ved at kontakte lektor Christian S. Jensen, Afdeling for Matematik og Datalogi, Aalborg Universitetscenter, Fredrik Bajers Vej 7, 9220 Aalborg Øst.

Af andet materiale om benchmarks kan anbefales "The benchmark Handbook for Database and Transaction Processing Systems," J Gray (editor), 2. udgave,

Morgan Kaufman, 1993. Denne samling af artikler beskriver en række benchmarks på forskellig vis. Desuden rummer den generelt materiale om konstruktion af benchmarks. Dernæst er "A Methodologi for Database System Performance Evaluation," af D. J. DeWitt og H. Boral (siderne 176–185 i "Proceedings of ACM SIGMOD '84") en god teoretisk artikel om nogle af de aspekter af databasesystemer, benchmarks bør teste.